ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Number Theory 104 (2004) 327–334
JOURNAL OF Number Theory
http://www.elsevier.com/locate/jnt

On an additive representation function

R. Balasubramanian* and Gyan Prakash

The Institute of Mathematical Sciences, CIT Campus, Taramani, Chennai 600113, India

Received 21 March 2003; revised 23 July 2003

Communicated by D. Goss

Abstract

Let A be an infinite subset of natural numbers, $n \in \mathbb{N}$ and X a positive real number. Let $r(n)$ denotes the number of solution of the equation $n = a_1 + a_2$ where $a_1 \leqslant a_2$ and $a_1, a_2 \in A$. Also let $|A(X)|$ denotes the number of natural numbers which are less than or equal to X and belong to A. For those A which satisfy the condition that for all sufficiently large natural numbers n we have $r(n) \neq 1$, we improve the lower bound of $|A(X)|$ given by Nicolas et. al. [NRS98]. The bound which we obtain is essentially best possible.

Keywords: Addition of sequence; Additive representation function

Let $\mathbb{N}$ denote the set of all natural numbers. If A is an infinite subset of $\mathbb{N}$ then we set

$$A(x) = \{a \leqslant x : a \in A\}.$$

Let $r(A, n)$ denote the number of solutions of the equation

$$n = a_i + a_j, \quad \text{where } a_i \leqslant a_j, \ \ a_i, a_j \in A.$$

Here and in what follows A will always denote an infinite subset of $\mathbb{N}$ such that there exists a natural number $n_0(A)$ such that

$$r(A, n) \neq 1 \quad \text{for } n \geqslant n_0(A).$$

*Corresponding author. Fax: +91-44-22541934.
E-mail addresses: balu@imsc.res.in (R. Balasubramanian), gyan@imsc.res.in (G. Prakash).

0022-314X/$ - see front matter
doi:10.1016/j.jnt.2003.09.004

Also $a_0(A)$ shall denote the least natural number such that $a_0(A) \in A$ and $a_0(A) \geqslant n_0(A)$. Regarding such sets, Nicolas et al. [NRS98] proved the following theorem:

Theorem. *If A is an infinite subset of $\mathbb{N}$ such that $r(A,n) \neq 1$ for all sufficiently large natural numbers n, then*

$$\limsup |A(x)| \left(\frac{\ln \ln x}{\ln x}\right)^{3/2} \geqslant \frac{1}{20}.$$

They also gave an example of a set A such that $r(A,n) \neq 1$ for all sufficiently large natural numbers n and $|A(x)| \ll (\ln x)^2$. In this paper we shall show the following:

Theorem 1. *There exists an absolute constant $c > 0$ with the following property: for any infinite subset A of $\mathbb{N}$ such that $r(A,n) \neq 1$ for all sufficiently large natural numbers n, then*

$$|A(x)| \geqslant c \left(\frac{\ln x}{\ln \ln x}\right)^2 \quad \textit{for all } x \textit{ sufficiently large.}$$

Theorem 1 follows from Proposition 2 by noting that if Y is sufficiently large then for some positive absolute constant c the interval $[(Y)^{\frac{1}{2}}, Y)$ contains at least $c(\frac{\ln Y}{\ln \ln Y})$ disjoint intervals of the form $[b, b(\ln Y)^{11})$.

Apart from the arguments used in proving Lemma 4, the rest of arguments used in this paper are as in [NRS98]. Lemma 4 improves inequality (1) of Proposition 1 and this improves result of Proposition 1 and gives Proposition 2. As remarked above Theorem 1 is an immediate corollary of Proposition 2. The sequence B_Y constructed in Lemma 3 is a slight modification of analogous sequence constructed in [NRS98] (see [NRS98, p. 304]).

Lemma 1. *For all real numbers $x > a_0(A)$ the interval $(x, 2x]$ contains an element of the set A.*

Proof. Let a be the largest element of A not exceeding x. Then $a \geqslant a_0(A)$ so that the integer $n = a + a$ is $> n_0(A)$. It now follows that there is a pair (c, d), with $c \leqslant d$, of elements of A distinct from the pair (a, a) such that $n = c + d$. Since $d \geqslant c$ this implies that $d > a$ whence $d > x$ by the choice of a. Clearly we have $d \leqslant n = a + a \leqslant 2x$. In summary, we have verified that the element d of A lies in $(x, 2x]$. $\square$

l-good interval: An interval $I = [k, k + l]$ is defined to be l-good if $I \cap A = \{k + l\}$; that is it is of length l, the last element is in A and no other element is in A.

Lemma 2. *Let Y be a sufficiently large real number and $|A(Y)| \leqslant (\ln Y)^2$. Then for any real number b such that $1 \leqslant b \leqslant \frac{Y}{2(\ln Y)^{10}}$ there exists b good interval in $[b(\ln Y)^5, 2b(\ln Y)^{10}]$.*

Proof. We consider interval $C = [b(\ln Y)^5, b(\ln Y)^{10}]$. Then the length of C is at least $\frac{1}{2}b(\ln Y)^{10}$ for all Y sufficiently large, but

$$|C \cap A| \leqslant |A(Y)| \leqslant (\ln Y)^2 < \tfrac{1}{2}(\ln Y)^{10}.$$

Therefore, there exists in C a closed interval I of length b and void of A. Moving I to right till it hits A, we get a b-good interval I'. Using Lemma 1 it follows that $I' \subset [b(\ln Y)^5, 2b(\ln Y)^{10}]$. $\square$

Lemma 3. *Let Y be a sufficiently large real number and $|A(Y)| \leqslant (\ln Y)^2$. Then there exists an increasing sequence $\{b_1, b_2, \dots, b_m\} = B_Y$ of elements of A not exceeding $\sqrt{Y}$ and satisfying the following properties:*

(I) *For each $1 \leqslant i \leqslant m-1, b_{i+1} \geqslant b_i(\ln Y)^5$.*
(II) *For each $1 \leqslant i \leqslant m-1, [b_{i+1} - b_i, b_{i+1})$ does not contain an element of the set A.*
(III) *The number of terms m of sequence B_Y is at least $c\frac{\ln Y}{\ln \ln Y}$ where c is a positive absolute constant.*

Proof. We shall define $B_Y = \{b_1, b_2, \dots, b_i, \dots\}$ recursively. We set $b_1 = a_0(A)$. Suppose $b_1, b_2, \dots, b_i$ have been determined and $b_i \leqslant \frac{1}{2}\frac{\sqrt{Y}}{(\ln Y)^{10}}$ then applying Lemma 2 we choose the smallest $a \in [b_i(\ln Y)^5, 2b_i(\ln Y)^{10}]$ such that $[a - b_i, a)$ does not contain any element of A. We set b_{i+1} to be a. The recursion is terminated if $b_i > \frac{1}{2}\frac{\sqrt{Y}}{(\ln Y)^{10}}$.

Let B_Y be a sequence constructed in manner described above. Clearly, (I) and (II) hold for each $1 \leqslant i \leqslant m-1$. Further for each i we have that

$$b_{i+1} \leqslant 2b_i(\ln Y)^{10} < b_i(\ln Y)^{11}$$

whence by induction $b_m < a_0(A)(\ln Y)^{11m}$. Since recursion terminates at b_m we have $b_m > \frac{1}{2}\frac{\sqrt{Y}}{(\ln Y)^{10}}$. These remarks imply (III). $\square$

In what follows, B_Y will denote the sequence constructed as in the proof of Lemma 3.

Proposition 1. *Let Y be a sufficiently large real number and $|A(Y)| \leqslant (\ln Y)^2$. Let b be any real number such that $[b, b(\ln Y)^{11}) \subset (\sqrt{Y}, Y)$. Then the number of elements of A*

contained in the interval $[b, b(\ln Y)^{11})$ *is* $> c(\frac{\ln Y}{\ln\ln Y})^{\frac{1}{2}}$ *where* c *is a positive absolute constant.*

Proof. Lemma 2 implies that there is an element a of the set A lying in the interval $[b, 3b(\ln Y)^{10})$ such that the interval $[a-b, a)$ does not contain any element of A. We choose one such a.

Let S denote the set of elements of A in the interval $[b, b(\ln Y)^{11})$ and s denote the cardinality of S. Let S_1 and S_2 denote the sets of elements of A in the intervals $[b, a)$ and $[a, b(\ln Y)^{11})$, respectively, and let s_1 and s_2 denote the cardinalities of S_1 and S_2, respectively. We then have $s = s_1 + s_2$.

For each i, $1 \leqslant i \leqslant m$, let $n_i = a + b_i$, where $\{b_i\} = B_Y$ is the sequence supplied by Lemma 3. Since each $n_i \geqslant Y^{1/2}$, we see that when Y is sufficiently large, each n_i is $\geqslant n_0(A)$. For each i we then choose a pair (c_i, d_i), with $d_i \geqslant c_i$, of elements of A distinct from the pair (a, b_i) such that $n_i = c_i + d_i$. For each i we then have either $d_i < a$ or $d_i > a$. Let P_1 denote the set of those pairs (c_i, d_i) with $d_i < a$ and P_2 the set of those pairs (c_i, d_i) with $d_i > a$. Let p_1 and p_2 denote the cardinalities of P_1 and P_2, respectively. We then have $p_1 + p_2 = m$.

If (c_i, d_i) is in P_1 we have $c_i \leqslant d_i < a - b$ and hence that $d_i \geqslant c_i = a + b_i - d_i \geqslant a - (a - b) = b$. In other words, c_i and d_i are elements of S_1. It follows that $S_1 \times S_1$ contains P_1. Consequently, we have that $s_1^2 \geqslant p_1$ or that

$$s_1 \geqslant p_1^{1/2}. \tag{1}$$

If (c_i, d_i) is in P_2 we have $a < d_i$. Further, we have that

$$c_i + d_i = a + b_i \leqslant 3b(\ln Y)^{10} + Y^{\frac{1}{2}} \leqslant 3b(\ln Y)^{10} + b \leqslant b(\ln Y)^{11} \tag{2}$$

and hence that $d_i \leqslant b(\ln Y)^{11}$. It follows that the mapping ϕ that associates (c_i, d_i) to d_i maps P_2 into S_2. Let us verify that ϕ is injective. Suppose to the contrary that (c_i, d_i) and (c_j, d_j) are elements of P_2 such that $d_i = d_j$ and $i < j$. Then

$$c_j \geqslant c_j - c_i = b_j - b_i \geqslant b_j - b_{j-1}. \tag{3}$$

Also $c_j < b_j$ because $c_j + d_j = a + b_j$ and $d_j > a$. It follows that the element c_j of A lies in the interval $[b_j - b_{j-1}, b_j)$ contradicting (ii) of Lemma 3. The injectivity of ϕ implies that $s_2 \geqslant p_2$.

In summary we have verified that

$$s = s_1 + s_2 \geqslant p_1^{1/2} + p_2 \geqslant p_1^{1/2} + p_2^{1/2} \geqslant (p_1 + p_2)^{1/2} \geqslant m^{1/2} \tag{4}$$

from which the proposition follows on recalling (III) of Lemma 3. □

Corollary. *There exists an absolute constant* $c > 0$ *with the following property*: *For any infinite subset* A *of* $\mathbb{N}$ *such that* $r(A, n) \neq 1$ *for all sufficiently large natural numbers* n, *we have*:

$$|A(Y)| \geqslant c\left(\frac{\ln Y}{\ln\ln Y}\right)^{\frac{3}{2}}.$$

Proof. The corollary follows from Proposition 1 on noting that if Y is sufficiently large then for some positive absolute constant c the interval $[(Y)^{\frac{1}{2}}, Y]$ contains at least $c\frac{\ln Y}{\ln\ln Y}$ disjoint intervals of the form $[b, b(\ln Y)^{11})$. □

Result in Proposition 1 can be improved and we have Proposition 2. Rest of arguments being the same, Proposition 2 follows by improving inequality (1) in Proposition 1 using Lemma 4. We shall first just state Lemma 4 and deduce Proposition 2. Later we shall prove Lemma 4 which require a few other lemmas.

Lemma 4. *With notations and assumptions as in Proposition* 1 *we have* $|[b,a) \cap A| \geqslant c|P_1|$, *where c is a positive absolute constant.*

Proposition 2. *Let Y be a sufficiently large real number and* $|A(Y)| \leqslant (\ln Y)^2$. *Let b be any real number* $\geqslant 1$ *such that* $[b, b(\ln Y)^{11}) \subset (\sqrt{Y}, Y)$. *Then the number of elements of A contained in the interval* $[b, b(\ln Y)^{11})$ *is* $> c(\frac{\ln Y}{\ln\ln Y})$ *where c is an positive absolute constant.*

Proof. Notice that assumptions of Propositions 1 are satisfied here. Then arguing as in proof of Proposition 1 and using Lemma 4 in place of inequality (1) Proposition follows. □

Lemmas 5 and 7 are required for proving Lemma 4.

Lemma 5. *Let* $B_Y = \{b_1, b_2, \ldots, b_m\}$ *be a sequence as constructed in Lemma* 3. *Suppose* $\sum_{i=1}^{n} x_i b_i = 0$ *where* $1 \leqslant n \leqslant m$ *and* $x_i \in \{1, -1, 0, 2, -2\}$ *for all* $1 \leqslant i \leqslant n$. *Then* $x_i = 0$ *for all i.*

Proof. Suppose it is not true and there exist sequence $\{x_i\}$ such that $\sum_{i=1}^{n} x_i b_i = 0$ where $1 \leqslant n \leqslant m$ and x_i is not zero for some i. Without loss of generality we may assume that $x_n \neq 0$. Then,

$$x_n b_n = \sum_{i=1}^{n-1} -x_i b_i.$$

As $n < m < |A(Y)| < (\ln Y)^2$ so $b_n \leqslant |x_n b_n| < 2(\ln Y)^2 b_{n-1}$. But by construction of B_Y, $b_n \geqslant (\ln Y)^5 b_{n-1}$. Hence there is a contradiction. □

Let us recall some definitions from graph theory which we need for our purpose. A graph G consists of a finite nonempty set $V = V(G)$ of vertices together with a prescribed set X of unordered pairs of elements of V. Each pair $x = \{u, v\}$ is an edge of G and is said to join u and v. Notice that a graph thus defined is a finite undirected graph without multiple edges but may have loops. A walk of a graph G is an alternating sequence of vertices and edges $v_1, x_1, v_2, \dots, v_{n-1}, x_{n-1}, v_n$, beginning and ending with vertices, in which each edge joins two vertices immediately preceding and following it. It is closed if $v_1 = v_n$. It is a trail if all the edges are distinct. By an *even closed trail* we shall mean a trail which is closed and have even number of edges. A cycle is a closed trail in which all the vertices are distinct. Two trails which define same subgraph are considered equivalent and are not distinguished.

Lemma 6. *Let G be a graph with no loops and no even closed trails. Then any two distinct closed trails in G are disjoint, that is, if T_1 and T_2 are two distinct closed trails in G, and $V(T_1)$ and $V(T_2)$ denote the set of vertices in T_1 and T_2, respectively, then $V(T_1) \cap V(T_2) = \emptyset$.*

Proof. Suppose it is not true. Then there exist two distinct closed trails T_1 and T_2 in G such that $V(T_1) \cap V(T_2) = V_c$ (say) $\neq \emptyset$. As T_1 and T_2 are two distinct trails so there is an edge in at least one of them which is not common to both of them. Let say x is one such edge and without loss of generality we may assume it is in T_1. Suppose $T_1 = v_1, x_1, v_2, x_2, v_3, \dots, v_i, x_i, v_{i+1}, \dots, v_{n-1}, x_n, v_n$. As $V_c \neq \emptyset$ so we may assume that $v_1 = v_n \in V_c$. Then if we choose $v_l \in V(T_2)$ nearest to left of x and $v_r \in V(T_2)$ nearest to right, in sequence for T_1 thus considered, then only vertices which $T = v_l, x_l, v_{l+1}, \dots, v_{r-1}, x_{r-1}, v_r$ share with T_2 are v_l and v_r. (It is possible that v_l is same as v_r.) Also then by choice of x, v_l, v_r the trail T does not have any common edge with T_2. As $v_l, v_r \in V(T_2)$ so there is a trail T'' in T_2 starting from v_l and ending with v_r. Now by choice of T we have that $T_u = T \cup T''$ is a closed trail. Also again by choice of T we have that $T_r = (T_2 \backslash T'') \cup T$ is another closed trail. (Notice that it may be so that $(T_2 \backslash T'')$ is empty but that does not affect our arguments.) Now it is clear that either T_u or T_r has an even number of edges depending on whether number of edges of T and T'' have same parity or different parity. But this is contrary to the assumption that G has no even closed trail. $\square$

Lemma 7. *Let G be a graph with n vertices and having no loops. Further assume that G has no even closed trail. Then number of edges in G, say $e(G)$, is at most $2n$.*

Proof. It is clearly enough to prove lemma in case when G is connected. From previous lemma no closed trail in G has a proper closed sub-trail. This implies that any closed trail is a cycle and any two cycles are disjoint. So G cannot have more than n cycles. Now, we shall show that $d(G) = e(G) -$ number of vertices is at most n and this proves the lemma. If we shrink all cycles in G to get new graph G' then G' has no cycle and is connected. So G' is a tree. But then $d(G') = -1$. Also as cycles in G are disjoint so $d(G) = d(G') +$ number of cycles in G. This implies that $d(G) \leqslant n - 1$. $\square$

Now we shall prove Lemma 4.

Lemma 4. *With notations and assumptions as in Proposition* 1 *we have*

$$|[b,a)\cap A| \geqslant c|P_1|,$$

where c is a positive absolute constant.

Proof. From Proposition 1 we recall that the set P_1 consists of pairs (c_j, d_j) of elements of the set A such that $c_j \leqslant d_j < a$. Also for each pair (c_j, d_j) belonging to the set P_1 there is exactly one term b_j of the sequence B_Y such that $c_j + d_j = a + b_j$. Let S_1 denote the set of elements of A lying in interval $[b, a)$, that is, $S_1 = [b,a)\cap A$. Then it was shown in Proposition 1 that $P_1 \subset S_1 \times S_1$.

We shall construct a graph G associated to the set P_1. As $P_1 \subset S_1 \times S_1$ we define $f_1 : P_1 \to S_1$ and $f_2 : P_1 \to S_1$ by $f_1(c_i, d_i) = c_i$, $f_2(c_i, d_i) = d_i$. The set of vertices of graph G, let say V, consists of those elements v of S_1 such that either v belongs to image of f_1 or of f_2. Then we have following upper bound on number of vertices of G.

$$|V| = n \leqslant |\text{Image of } f_1| + |\text{Image of } f_2| \leqslant 2|S_1| = 2|[b,a)\cap A|. \tag{5}$$

The set of edges of G (say X) consists of those unordered pair $\{v_1, v_2\}$ of V such that either (v_1, v_2) or $(v_2, v_1) \in P_1$. In other words two vertices v_1 and v_2 are joined by an edge if and only if either (v_1, v_2) or $(v_2, v_1) \in P_1$. The graph G thus constructed satisfy following properties:

(I) There is a natural one–one correspondence between edges of G and elements of P_1.
(II) If x is an edge in G joining vertices v_1 and v_2 then there is a term b_x in the sequence B_Y such that $v_1 + v_2 = a + b_x$.
(III) For two distinct edges x and y, the corresponding b_x and b_y given as above are distinct.

All these properties are easily verified using definition of G and P_1. So (I) in particular implies that number of edges in G is same as number of elements in P_1. Then to prove the lemma it is enough to show that

$$\text{number of edges in } G = e(G) \leqslant cn \text{ for some positive absolute constant } c. \tag{6}$$

Now G can have at most n loops. So if we remove all loops from G to get another graph G_1 then to show (6) it is enough to show that

$$e(G_1) \leqslant cn \text{ for some positive absolute constant } c.$$

We claim that G_1 does not have any even closed trail. Then using claim and Lemma 7 we have (6).

Suppose claim is not true and G_1 has an even closed trail

$$T = v_1, x_1, v_2, x_2, v_3, \ldots, v_i, x_i, v_{i+1}, \ldots, v_{2m-1}, x_{2m-1}, v_{2m}, x_{2m}, v_1,$$

where v_i is a vertex of G and x_i is an edge joining vertices immediately preceding and following it. Also by definition of trail we have, for $1 \leqslant i, j \leqslant 2m$ and $i \neq j$, $x_i \neq x_j$. Then using property (II) of G we have

$$v_i + v_{i+1} = a + b_i, \quad \text{where } 1 \leqslant i \leqslant 2m-1,$$

$$v_{2m} + v_1 = a + b_{2m},$$

where $b_i \in \{b_i\}$ for all $1 \leqslant i \leqslant 2m$. Further using property (III) of G it follows that for $1 \leqslant i, j \leqslant 2m$ and $i \neq j$ we have $b_i \neq b_j$. Now we have

$$\sum_{i=1}^{2m-1} (-1)^i (v_i + v_{i+1}) = \sum_{i=1}^{2m-1} (-1)^i (a + b_i), \tag{7}$$

$$v_{2m} + v_1 = a + b_{2m}. \tag{8}$$

Adding (7) and (8) we get

$$0 = \sum_{i=1}^{2m} (-1)^i b_i$$

which is a contradiction to Lemma 5. $\square$

References

[NRS98] J.-L. Nicolas, I.Z. Ruzsa, A. Sárközy, On the parity of additive representation functions, J. Number Theory 73 (1998) 292–317.